CHAPTER IV : ADMISSIBILITY OF MURTHY'S AND MIDZUNO ESTIMATORS IN THE CLASS OF LINEAR UNBIASED ESTIMATORS

4.0 SUMMARY

Murthy [10], [11] proposed two estimators for the total of a finite population. Joshi [8] indicated a method for proving the admissibility of these estimators. However, we show here that Joshi's method is not applicable. The admissibility of Murthy's estimators is established by a modification of the method of chapter III. The same modification also leads to a proof of the admissibility of some estimators for the Midzuno sampling scheme [9].

4.1 GENERAL RESULT

As remarked in Chapter III, the admissible linear invariant and unbiased estimator of theorem 3.4.1 is difficult to obtain in practice. Hence we modify the method slightly as follows :

Let $k_1, k_2, \ldots, k_N$ be non-zero constants such that $\sum_{i=1}^{N} k_i = 1$. Let $q_1, q_2, \ldots, q_N$ be positive numbers. Then we try to minimise $\sum_{i=1}^{N} q_i V_i$ subject to conditions (3.1.1) of unbiasedness and zero variance at the point $(k_1, \ldots, k_N)$.

The estimator $t(s, \underset{\sim}{Y})$ of (1.2.6) will have zero variance at the point $(k_1, \ldots, k_N)$ if

$$\sum_{i \in s} b(s,i)\, k_i = 1, \qquad s \in S \qquad \ldots(4.1.1)$$

and $t(s, \underset{\sim}{Y})$ is unbiased if (3.1.1) hold. Therefore to minimise $\sum_{i=1}^{N} q_i V_i$ subject to (3.1.1) and (4.1.1), we consider

$$\phi = \sum_{i=1}^{N} q_i V_i - 2 \sum_{i=1}^{N} \lambda_i \sum_{s \ni i} b(s,i) p(s)$$
$$-2 \sum_{s \in S} \mu_s \sum_{i \in s} b(s,i)\, k_i,$$

where λ_i and μ_s denote Lagrange's multipliers. Equating the partial derivatives of ϕ w.r.t. $b(s,i)$ to zero, we get

$$q_i b(s,i)\, p(s) = \lambda_i p(s) + \mu_s k_i . \qquad \ldots(4.1.2)$$

Write $\xi_i = \lambda_i / q_i$, $a_s = \mu_s / p(s)$ and $\eta_i = k_i / q_i$.

Then (4.1.2) reduces to

$$b(s,i) = \xi_i + a_s \eta_i . \qquad \ldots(4.1.3)$$

Conditions (4.1.1), together with (4.1.3), give

$$a_s = (1- \sum_{j \in s} \xi_j k_j)/ \delta(s), \qquad \ldots(4.1.4)$$

where $\delta(s) = \sum_{j \in s} \eta_j k_j$. Condition (3.1.1) now gives

$$\xi_i \pi_i - \eta_i \sum_{s \ni i} \frac{p(s)}{\delta(s)} \sum_{j \in s} \xi_j k_j = 1 - \eta_i \sum_{s \ni i} \frac{p(s)}{\delta(s)}, \qquad \ldots(4.1.5)$$

where $i = 1, 2, \ldots, N$ and π_i denotes the inclusion probability for the unit i. After routine algebraic simplification (4.1.5) can be written as

$$A\underset{\sim}{\xi} = \underset{\sim}{d}, \qquad \ldots(4.1.6)$$

where the matrix A and the column vector $\underset{\sim}{d}$ are given by

$$a_{ii} = \pi_i - \eta_i k_i \sum_{s \ni i} [p(s)/ \delta(s)],$$

$$a_{ij} = - \eta_i k_j \sum_{s \supset \{i,j\}} [p(s)/ \delta(s)], \qquad i \neq j,$$

and

$$d_i = 1 - \eta_i \sum_{s \ni i} [p(s)/ \delta(s)].$$

Given a solution $\underset{\sim}{\xi}$ of (4.1.6) we can use (4.1.4) and (4.1.3) to write b(s,i) as

$$b(s,i) = \xi_i + [\eta_i (1- \sum_{j \in s} \xi_j k_j)/ \delta(s)]. \qquad \ldots(4.1.7)$$

To claim the admissibility of the resulting estimator we have to show that the quantities $b(s,i)$ of (4.1.7) are independent of the choice of the particular solution $\underset{\sim}{\xi}$ of (4.1.6).

Theorem 4.1.1 : Suppose the matrix A has rank (N-1). Then the system (4.1.6) is consistent and the $b(s,i)$ computed from (4.1.7) is the same for all solution $\underset{\sim}{\xi}$ of (4.1.6). The resulting estimator is admissible within the class of all unbiased linear estimators of the population total.

Proof : Let S_i denote the set of those samples s which contain the unit i. Let $\underset{\sim}{k} = (k_1, \ldots, k_N)'$. The i^{th} component of $\underset{\sim}{k}'A$ is

$$k_i\pi_i - k_i \eta_i k_i \sum_{s \ni i} [p(s)/\delta(s)] - \sum_{j\neq i} \eta_j k_j k_i \sum_{s \supset \{i,j\}} [p(s)/\delta(s)]$$

$$= k_i\pi_i - \sum_{j=1}^{N} \sum_{s\supset\{i,j\}} \eta_j k_j k_i [p(s)/\delta(s)]$$

$$= k_i\pi_i - \sum_{s \in S_i} \sum_{j \in s} k_i \eta_j k_j [p(s)/\delta(s)]$$

$$= k_i\pi_i - \sum_{s \in S_i} \frac{p(s)}{\delta(s)} k_i \sum_{j \in s} \eta_j k_j$$

$$= k_i\pi_i - \sum_{s \ni i} k_i p(s) = 0.$$

Since rank of A is N-1 and $\underset{\sim}{k}'A = \underset{\sim}{0}$, the system (4.1.6) is consistent as soon as $\underset{\sim}{k}'\underset{\sim}{d} = 0$.

Now
$$\underset{\sim}{k}'\underset{\sim}{d} = \sum_{i=1}^{N} k_i d_i = \sum_{i=1}^{N} k_i \left[1-\eta_i \sum_{s \ni i} \frac{p(s)}{\delta(s)}\right]$$
$$= 1 - \sum_{s \in S} \frac{p(s)}{\delta(s)} \cdot \sum_{i \in s} \eta_i k_i$$
$$= 1 - \sum_{s \in S} p(s) = 0.$$

Thus (4.1.6) is consistent.

The vector $\underset{\sim}{\eta} = (\eta_1, \ldots, \eta_N)'$ is such that

$A\underset{\sim}{\eta} = \underset{\sim}{0}$. Therefore a general solution of (4.1.6) is $\underset{\sim}{\xi} = \underset{\sim}{\xi}^* + C\underset{\sim}{\eta}$ where $\underset{\sim}{\xi}^*$ is a particular solution and C is arbitrary. Now it is easy to check that b(s,i) computed from (4.1.7) does not depend on C. Thus the estimator given by (4.1.7) is the unique estimator which minimizes $\sum_{i=1}^{N} q_i V_i$ amongst all unbiased linear estimators which attain zero variance at $\underset{\sim}{k}$. Therefore the estimator given by (4.1.7) is admissible in the linear unbiased class. This completes the proof of the theorem.

In Chapter III we defined the term 'connected sampling design'. This is used in the following theorem.

<u>Theorem</u> 4.1.2 : <u>If the sampling design is connected then rank of</u> A is N-1

Where A is the coefficient matrix in the equations (4.1.6).

<u>Proof</u> : Since $k_i \neq 0$ and $q_i > 0$, we have

$\eta_i = k_i/q_i \neq 0$, and $\eta_i k_i = k_i^2/q_i > 0$. Thus $\delta(s) > 0$.

Let $\Gamma = \text{diag}(k_1, \ldots, k_N)$; $\Omega = \text{diag}(\eta_1, \ldots, \eta_N)$ and $C = \Gamma A \Omega$.

The entries of C are given by

$$C_{ii} = \pi_i k_i \eta_i - k_i^2 \eta_i^2 \sum_{s \ni i} [p(s)/\delta(s)] \quad \text{and}$$

$$C_{ij} = -k_i \eta_i k_j \eta_j \sum_{s \supset \{i,j\}} [p(s)/\delta(s)], \quad i \neq j.$$

Since $\eta_i k_i > 0$ and $\delta(s) > 0$, it follows that $C_{ij} \leqslant 0$ for $i \neq j$. Further it can be easily checked that $\sum_{j=1}^{N} C_{ij} = 0$ for all i. Finally C is symmetric. Thus C satisfies the three properties used in the proof of theorem 3.2.1. We can therefore conclude that rank of $C = (N-1)$ whenever the sampling design is connected. But A and C have the same rank as Γ and Ω are non-singular. The proof of the theorem is thus complete.

Sampling designs used in survey sampling are usually connected. Therefore the applicability of the theory of this chapter depends upon the easy solvability of (4.1.6). In the next two sections we show that the theory leads to the proofs of the admissibility of some known estimators.

4.2 ADMISSIBILITY OF MURTHY'S ESTIMATORS

Suppose we take a sample of 2 units without replacement from $\mathcal{U}$ with probabilities proportional to $p_1, \ldots, p_N$ at each stage. Here $p_i > 0$ for all i and $\sum_{i=1}^{N} p_i = 1$. Suppose the sample obtained is $s = \{i,j\}$. Then Murthy [10], [11] proposed the following two estimators for the population total.

$$T_1(s) = \frac{1}{2-p_i-p_j}\left[\frac{(1-p_j)\,Y_i}{p_i} + \frac{(1-p_i)\,Y_j}{p_j}\right] \quad \ldots(4.2.1)$$

and

$$T_2(s) = \frac{1}{2-p_i-p_j}\left[\frac{(1-p_i)\,Y_i}{\pi_i - p_i} + \frac{(1-p_j)\,Y_j}{\pi_j - p_j}\right]. \quad \ldots(4.2.2)$$

Joshi [8], after proving the admissibility of the Sen-Yates-Grundy variance estimator for designs of fixed sample size 2, claimed that the same proof can be made applicable to the estimators T_1 and T_2. This claim is, however, incorrect. The reason is that Joshi's[8] equation (28) depends on the fact that $v(s, x^{iii}) = v(s, x^{iv})$, which is true for the SYG estimator but false for the estimators T_1 and T_2. As the following example shows the criterion in section 4 of Joshi's paper is incorrect.

Example : Suppose that each of the samples $s_i = \{i, i+1\}$ has probability $1/N$. Here we interprete $(i+1)$ as 1 when $i=N$. Define the estimator T_α by $T_\alpha(s_i)=N[\alpha Y_i+(1-\alpha)Y_{i+1}]$. Then T_α is unbiased for the population total. Further, if $Y_i = k$ for all i, then $T_\alpha(s_i)$ Nk for all i. However, T_α is inadmissible whenever $\alpha \neq \frac{1}{2}$, whereas each T_α should be admissible according to Joshi's criterion.

We now proceed to prove the admissibility of T_1 and T_2. In the general theory of section 4.1, take $q_i = k_i^2/(1-p_i)$. Then

$$\eta_i = k_i/q_i = (1-p_i)/k_i$$

and $\eta_i k_i = (1-p_i)$. Therefore, if $s = \{i, j\}$ then

$\delta(s) = 2-p_i-p_j$. On the other hand

$$p(s) = p_i p_j(2-p_i-p_j)/\left[(1-p_i)(1-p_j)\right] ,$$

and

$$\pi_i = p_i + p_i \sum_{j\neq i} \left[p_j/(1-p_j)\right]$$

Therefore $\frac{p(s)}{\delta(s)} = \frac{p_i p_j}{(1-p_i)(1-p_j)}$

and $a_{ii} = \pi_i - \eta_i k_i \sum_{s \ni i} [p(s)/\delta(s)] = \pi_i - p_i \sum_{j \neq i} [p_j/(1-p_j)]$

$= p_i$.

Further, for $i \neq j$

$$a_{ij} = -\eta_i k_j . \frac{p(s)}{\delta(s)} = -\frac{p_i p_j k_j}{k_i(1-p_j)} .$$

Finally $d_i = 1 - \frac{p_i}{k_i} \sum_{j \neq i} [p_j/(1-p_j)]$.

The system (4.1.6) can thus be written as

$$p_i \xi_i - \frac{p_i}{k_i} \sum_{j \neq i} \frac{p_j k_j}{(1-p_j)} \xi_j = 1 - \frac{p_i}{k_i} \sum_{j \neq i} \frac{p_j}{(1-p_j)}, \quad \ldots(4.2.3)$$

where i=1,2, ..., N. Fortunately, (4.2.3) can be solved explicitly. A solution is

$$\xi_i = \frac{1}{k_i} + \frac{(1-p_i)}{p_i}, \quad i=1,2, \ldots, N. \quad \ldots(4.2.4)$$

To verify this observe that with ξ_i as in (4.2.4), the left side of (4.2.3) becomes

$$\frac{p_i}{k_i} + (1-p_i) - \frac{p_i}{k_i} \sum_{j \neq i} \frac{p_j + (1-p_j)k_j}{(1-p_j)}$$

$$= \frac{p_i}{k_i} + (1-p_i) - \frac{p_i}{k_i} \sum_{j \neq i} [p_j/(1-p_j)] - \frac{p_i}{k_i} \sum_{j \neq i} k_j .$$

Note that $\sum_{j \neq i} k_j = 1 - k_i$. Therefore the last expression equals

$$\frac{p_i}{k_i} + (1-p_i) - \frac{p_i}{k_i} \sum_{j\neq i} [p_j/(1-p_j)] - \frac{p_i}{k_i}(1-k_i)$$

$$= 1 - \frac{p_i}{k_i} \sum_{j\neq i} [p_j/(1-p_j)] \ .$$

This is the right side of (4.2.3). Given the solution (4.2.4), the quantity b(s, i) can be computed from (4.1.7). Routine algebra yields :

$$b(s,i) = \frac{1-p_j}{k_i(2-p_i-p_j)} + \frac{k_i p_j - k_j p_i}{k_i p(s)} \ . \qquad \ldots(4.2.5)$$

Since the design is connected, (4.2.5) gives a large class of admissible linear estimators for the particular sampling scheme. Special cases yield Murthy's estimators.

Case 1. Let $k_i = p_i$. Observe that the second term on the right side of (4.2.5) then reduces to zero. Therefore

$$b(s,i) = (1-p_j)/\left[p_i(2-p_i-p_j)\right] \ .$$

The resulting estimator is identical with Murthy's estimator T_1 in (4.2.1).

Case 2. Let $k_i = \pi_i - p_i$. Then

$$\sum_{i=1}^{N} (\pi_i - p_i) = \sum_{i=1}^{N} \pi_i - \sum_{i=1}^{N} p_i = 2-1 = 1, \text{ as required.}$$

We may use the general formula (4.2.5). However, observe that $d_i = 1 - \dfrac{\pi_i - p_i}{k_i}$. Therefore $d_i = 0$ for all i, in the present case. In other words we are dealing with the homogeneous system. Therefore, we may use the solution $\underset{\sim}{\xi} = \underset{\sim}{0}$. We then get, from (4.1.7),

$$b(s,i) = \frac{\eta_i}{\delta(s)} = \frac{(1-p_i)}{(2-p_i-p_j)(\pi_i-p_i)} \ .$$

The resulting estimator is identical with Murthy's estimator T_2 in (4.2.2).

<u>Case 3</u>. To construct a linear invariant estimator, we take $k_i = 1/N$. We then get

$$b(s,i) = \frac{N(1-p_j)}{(2-p_i-p_j)} + \frac{(p_j-p_i)}{p(s)} \quad \text{from (4.2.5).} \qquad \ldots(4.2.6)$$

<u>Case 4</u>. Let $k_i = p_i/\left[\alpha(1-p_i)\right]$ where α is so chosen that $\sum_{i=1}^{N} k_i = 1$. In this case

$$b(s,i) = \frac{(1-p_i)\left[\alpha(1-p_j) + (p_i-p_j)\right]}{p_i(2-p_i-p_j)} \ . \qquad \ldots(4.2.7)$$

The estimators defined by (4.2.6) and (4.2.7) seem to be new.

4.3 APPLICATION TO MIDZUNO'S SCHEME

Midzuno [9] proposed a sampling scheme under which the first unit is drawn with probabilities proportional to $p_1, \ldots, p_N$ and the remaining $(n-1)$ sample units are drawn from remaining $(N-1)$ population units with equal probability and without replacement. Here also, each $p_i > 0$ and $\sum_{i=1}^{N} p_i = 1$. Here S is clearly the class of all subsets of $\mathcal{U}$ of size n. Further

$$p(s) = \sum_{i \in s} p_i / \binom{N-1}{n-1}, \quad s \in S.$$

In the theory of section 4.1, take $q_i = k_i^2/p_i$. Then $\eta_i = p_i/k_i$ and $\delta(s) = \sum_{i \in s} p_i$. Therefore

$$\frac{p(s)}{\delta(s)} = \left\{ 1 / \binom{N-1}{n-1} \right\},$$

$$\sum_{s \ni i} \frac{p(s)}{\delta(s)} = 1 \quad \text{and}$$

$$\sum_{s \supset \{i,j\}} \left[p(s)/\delta(s) \right] = \left\{ \binom{N-2}{n-2} / \binom{N-1}{n-1} \right\} = \frac{n-1}{N-1}.$$

Moreover,

$$\pi_i = p_i + \left[(1-p_i)\,(n-1)/(N-1) \right].$$

Therefore,

$$a_{ii} = \pi_i - p_i = (1-p_i)(n-1)/(N-1),$$

and, for $i \neq j$

$$a_{ij} = p_i k_j (n-1)/[(N-1)k_i] \ .$$

The system (4.1.6) thus reduces to

$$(1-p_i)\xi_i - \frac{p_i}{k_i} \sum_{j \neq i} k_j \xi_j = \frac{N-1}{(n-1)} (1 - \frac{p_i}{k_i}). \quad \ldots(4.3.1)$$

Luckily, (4.3.1) also can be solved explicitly.

A solution is given by

$\xi_i = (N-1)/(n-1)$ for all i.

With this solution, (4.1.7) gives

$$b(s,i) = \frac{p_i}{k_i \sum_{j \in s} p_j} + \frac{N-1}{n-1} \quad \frac{k_i \sum_{j \in s} p_j - p_i \sum_{j \in s} k_j}{k_i \sum_{j \in s} p_j} \ . \quad \ldots(4.3.2)$$

Since the design is connected, (4.3.2) gives a class of admissible linear unbiased estimators for the Midzuno scheme. We will mention some special cases.

Case 1. Let $k_i = p_i$, Then the second term in (4.3.2) drops out. Therefore

$b(s,i) = \dfrac{1}{\sum_{j \in s} p_j}$. The resulting estimator can

be written as

$$T = \sum_{i \in s} Y_i / [\sum_{i \in s} p_i],$$

which is the usual unbiased ratio estimator for this scheme.

Case 2. Let $k_i = (1-p_i)/(N-1)$.

Routine algebra yields

$$b(s,i) = \frac{N-1}{(n-1)(1-p_i)} \left[1 - \frac{p_i}{\sum_{j \in s} p_j} \right]. \qquad \ldots(4.3.3)$$

If n=2 and $s = \{i,j\}$, then the estimator in (4.3.3) can be written as

$$T = \frac{(N-1)}{p_i + p_j} \left[\frac{p_j Y_i}{(1-p_i)} + \frac{p_i Y_j}{(1-p_j)} \right]. \qquad \ldots(4.3.4)$$

The estimator (4.3.4) has been given by Murthy [11].

Case 3. To construct a linear invariant estimator let $k_i = \frac{1}{N}$. After some simplification we get

$$b(s,i) = \frac{N}{n} + \frac{N-n}{n-1} \left(\frac{1}{n} - \frac{p_i}{\sum_{j \in s} p_j} \right)$$

$$= \frac{N-1}{n-1} - \frac{N-n}{n-1} \frac{p_i}{\sum_{j \in s} p_j}. \qquad \ldots(4.3.5)$$

Remark : In Chapter III we had $k_i = N^{-1}$ and $q_i=1$ for all i. Thus relationship between k_i and q_i did not depend on the sampling design. Hence we got an intractable system of equations. In this chapter we have $q_i = \frac{k_i^2}{(1-p_i)}$ in section 4.2 and $q_i = \frac{k_i^2}{p_i}$ in section 4.3. Thus the sampling design influences the relationship between q_i and k_i. This seems to be the main reason for our being able to solve (4.2.3) and (4.3.1). In particular we have been able to obtain explicit admissible linear invariant estimators : (4.2.6); (4.3.5).